ARI Research Note 95-27

Reducing the Confirmation Bias in an Evolving Situation

Martin A. Tolcott and F. Freeman Marvin

Decision Science Consortium, Inc.

Research and Advanced Concepts Office
Michael Drillings, Acting Director

February 1995

United States Army
Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Director

Research accomplished under contract for the Department of the Army

Decision Science Consortium, Inc.

Technical review by

Guy Siebold

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1995, February	Interim Sep 87 - Aug 88

4. TITLE AND SUBTITLE
Reducing the Confirmation Bias in an Evolving Situation

5. FUNDING NUMBERS
MDA903-86-C-0332
601102A
B74F
1901
C30

6. AUTHOR(S)
Tolcott, Martin A.; Marvin, F. Freeman (Decision Science Consortium, Inc.)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Decision Science Consortium, Inc.
1895 Preston White Drive, Suite 300
Reston, VA 22091

8. PERFORMING ORGANIZATION REPORT NUMBER
TR 88-11

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
ATTN: PERI-BR
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARI Research Note 95-27

11. SUPPLEMENTARY NOTES
Contracting Officer's Representative, Michael Drillings.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release;
distribution is unlimited.

12b. DISTRIBUTION CODE
--

13. ABSTRACT *(Maximum 200 words)*

This report presents the results of Phase 2 research on decision making in an evolving situation. As in Phase 1, the problem context was situation assessment by trained Army intelligence analysts working in pairs. Participants were given an initial battlefield scenario and asked to determine the enemy's most likely avenue of approach and to give their level of confidence; subsequently they were asked to reconsider their decisions after receiving each of three updated intelligence reports that contained some items that confirmed and some that did not confirm their early hypothesis. Finally, the participants were asked to rate each information item in terms of the degree to which it supported or contradicted their hypotheses.

14. SUBJECT TERMS
Decision
Judgment
Situation assessment
Information handling
Evolving decisions

15. NUMBER OF PAGES
41

16. PRICE CODE
--

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

ACKNOWLEDGEMENTS

We are grateful for the support and encouragement of Dr. Michael Drillings, Basic Research Office, Army Research Institute.

The research would not have been possible without the cooperation of Dr. Julie Hopson, Chief, ARI Field Unit, Fort Huachuca, Arizona, and of LCol Lamond and Captain Graham, of the U.S. Army Intelligence Center and School at Fort Huachuca. We are grateful to them and to the students and instructors who served as participants in the exercise.

We also gratefully acknowledge the help of Terry A. Bresnick in developing the problem materials, and Paul E. Lehner in the data analysis.

EXECUTIVE SUMMARY

This report presents the results of Phase 2 of research on decision making in an evolving situation. As in Phase 1, the problem context was situation assessment by trained Army intelligence analysts working in pairs. Participants were given an initial battlefield scenario, asked to determine the enemy's most likely avenue of approach, and give their level of confidence; subsequently they were asked to reconsider their decisions after receiving each of three updated intelligence reports, which contained some items that confirmed and some that disconfirmed their early hypothesis. Finally they were asked to rate each information item in terms of the degree to which it supported or contradicted their hypothesis.

In Phase 1 it was found that regardless of the initial hypothesis, confidence was high and tended to increase as the situation evolved. Confirming evidence was weighted significantly higher than disconfirming evidence, supporting other research in different contexts. Only one pair of participants (out of 11) changed their initial hypothesis. Graphic rather than analytic approaches were typical, and base rates were largely ignored in resolving uncertainties.

In Phase 2, participants were given a brief description of typical decision biases and of the Phase 1 findings, and were provided with graphic aids to facilitate their handling of uncertainties and to foster their awareness of alternative hypotheses. Results indicated a generally lower level of confidence, greater consideration of alternative enemy courses of action, and much more willingness to reverse early decisions based on new evidence. Half the teams (5 out of 10) changed their hypothesis at least once during the exercise. The tendency to overweight the importance of confirming evidence, although not eliminated, was significantly reduced.

The Phase 1 findings should showed that trained personnel, working on problems in their area of expertise, can show tendencies toward confirmation of early decisions and other non-normative cognitive behaviors similar to those found in less realistic laboratory tasks. In an evolving situation their interpretation of new information can be influenced by models or schemata based on early situation assessment. The findings of Phase 2 indicate that these tendencies can be reduced, although not entirely eliminated, by training innovations and by graphic aids that foster an awareness of uncertainty and provide help in dealing with it.

TABLE OF CONTENTS

FIGURES

TABLES

TABLES (cont'd)

1.0 INTRODUCTION

1.1 Objective

The overall objective of this research is to extend our theoretical understanding of decision making in situations that evolve over time, where the decision maker is presented with new or continually changing information. In the specific research reported here, interest focuses on the effect of an early decision on the handling of new information, some of which supports and some of which contradicts that decision. Of particular interest is the extent to which confidence in the early decision is maintained in the face of new information, and the extent to which confirmatory as compared with contradictory evidence is sought and attended to. The research also examines the extent to which the results of academic research using artificial problems are found to hold with trained personnel performing their occupational specialty, in this case, Army military intelligence analysts estimating an enemy's most likely avenue of approach. Finally, in this second phase of the research, several decision aiding and instructional techniques are tested to determine if they are effective in counteracting the biases found during the first phase.

1.2 Background

In Phase 1 of this research (Tolcott, Marvin and Lehner, 1987), Army intelligence analysts, working in pairs, were given a realistic (European) battlefield scenario with maps, overlays and Workbooks. They were given a description of recent events and the current situation, and were asked to make preliminary decisions about the most likely enemy avenue of approach, and to give their level of confidence. Subsequently, they were given updated intelligence reports containing some items that confirmed their early decision, some that contradicted it, and some that were neutral, and they were asked to reconsider their early decision and their confidence level, and revise them if appropriate. Three such updating trials were presented. Finally, they were asked to review all items in the update reports, and rate each item in terms of the degree to which it confirmed or contradicted their

initial hypothesis. One-third of the participants were given an initial scenario that slightly favored a northern avenue of approach, one-third a southern approach, and one-third a balanced or central approach. All participants received identical update reports.

The results showed that regardless of the initial hypothesis, confidence was generally high and tended to increase as the situation evolved. Confirming information was weighted significantly higher than disconfirming information. Contradictory evidence was usually recognized as contradictory, but was regarded as lower in importance; it was often judged as neutral, and sometimes as deliberately deceptive. Analysts with more experience would predict confirmatory events; their occurrence had a strong positive effect, while their non-occurrence led to later lowering of confidence. However, only one pair of participants (out of 11) changed their initial decision.

Familiar ("available") classes of information played a large role in decisions. Graphic/intuitive approaches (judgments based on information plotted on maps and overlays) were more common than tabular/analytic ones (reference to data in the Workbooks), and base rates were largely ignored in resolving uncertainties. The analysts appeared to model the situation based on early information, and to be influenced by new information according to how consistent it was with their model.

This research is part of a growing body of behavioral decision research showing that people depart from normative models when making judgments under conditions of uncertainty. For example, Gettys and Fisher (1979) and Gettys et al. (1981) have shown that people are poor at generating hypotheses and options for action. Wason (1960), Einhorn (1980), Einhorn and Hogarth (1978), and others have shown that people tend to stubbornly hold to a hypothesis generated early, ignore disconfirming evidence, and in fact seek confirming evidence. Lopes (1981) has proposed that people integrate new information into old judgments by adjusting the old judgment toward the new information, producing a new value that lies somewhere between the two, by an "anchoring and adjustment" process (rather than by using a Bayesian model).

More recently, Einhorn and Hogarth (1987) have proposed a model of belief updating that incorporates not only the anchoring-and-adjustment principle,

but also the evaluation of evidence as confirming or disconfirming a hypothesis, and the conflicting tendencies of adaptation and inertia. In Einhorn and Hogarth's model, the degree of adaptation of a prior belief to new evidence depends on the strength of the belief and the direction of the evidence; the stronger the belief, the more it is discounted by negative evidence and the less it is strengthened by positive evidence. Thus, when evidence is mixed, the sequence in which it is presented plays a critical role in the strength of the final belief. The effect of inertia (reluctance to change) takes two forms: one is a tendency for inertia to increase over time (an increasing primacy effect), while the other is a constant level that depends on individual attitudes toward negative and positive evidence. Einhorn and Hogarth's data, obtained from university students performing a variety of paper-and-pencil tasks, generally support the model. They also suggest that, for tasks that are relatively low in cognitive domains, inertia is greater when the updated belief is called for after all the new evidence has been presented, than if new responses are requested after each piece of evidence is received (step-by-step).

In a recent study, Serfaty, et al. (1988) investigated the effects of new information on both prior beliefs and on planning based on these beliefs. For half their subjects they used essentially the same scenario that we did in our Phase 1 experiment; for the other half they used a Korean scenario. After each update report their subjects were asked to estimate the most likely enemy attack direction, assess its probability, give their confidence in the information to date, and adjust own troop positions if appropriate. One major difference between their study and ours is that they systematically varied the sequence in which confirming and disconfirming evidence was presented, in order to test the Einhorn and Hogarth model. Another major difference was that their update reports definitely either confirmed or disconfirmed the previous judgment, whereas in our case, each update report contained both confirming and disconfirming items (as well as neutral ones).

In general, Serfaty, et al.'s results tended to support the Einhorn and Hogarth model. Their subjects typically changed their hypothesis probabilities in the direction of the incoming information, showing strong recency effects, and the Einhorn and Hogarth model predicted the actual data slightly better than a Bayesian model did. The direction in which confidence

in the reliability of the information changed depended on the consistency of the source; thus, two successive confirming or disconfirming reports raised the confidence level, while a confirming followed by a disconfirming report, or the reverse, led to a decrease in confidence. However, the authors reported a remarkable stability in the level of confidence across conditions. One of the most interesting findings was that in the assignment and reassignment of troops (the planning as compared with the situation assessment activity), the subjects hedged by preparing for the two likely avenues of attack, rather than committing themselves to a single hypothesis, and changes to the initial assignments were minimal, rather than in accordance with changes in their beliefs. This is consistent with observations we made during exercises at Fort Carson in December 1987, where only minor adjustments to an initial plan were made regardless of changes in intelligence information being received.

In this connection, it should be noted that part of the motivation for the work by Serfaty, et al. was to examine the Headquarters Effectiveness Assessment Tool (HEAT), designed to measure commanders' decision making processes; similarly, the exercises we observed at Fort Carson were, in part, designed to assess the Army Command and Control Evaluation System (ACCES), a follow-on to HEAT. In both of these performance measurement systems, a premium is put on the extent to which a commander's initial plan can handle a change in the enemy's intent, without requiring radical modification of troop assignments. Thus commanders are, in effect, marked down if they revise their plans significantly, and are credited if their initial plans need only minor revision. This scoring system is an invitation to conservative or hedging action. Of course, as a practical matter the radical shifting of troops on the battlefield is a time-consuming and costly action, and the dynamics of movement is a good reason for avoiding it if possible. The issue may be formulated in terms of the distinction between planning and actual commitment. One might argue that during the planning phase, prior to commitment, flexibility should be emphasized and innovation encouraged, while after troops have been committed, minimum changes in plan should be emphasized. The distinction between these two phases of operations needs to be addressed, both in research on the decision process and in the development of battlefield evaluation tools.

The findings of our Phase 1 experiment, and those of Serfaty, et al. have important implications not only for the theoretical investigation of evolving decisions, but for the practical aspects of military intelligence analysis and operational planning. If the response to new information is influenced by earlier judgments, or by the sequence in which the information is received, then inferences about enemy intent will be affected by irrelevant considerations. Furthermore, the existence of a confirmation bias in situation assessment may have a significantly detrimental effect on the management of intelligence information collection, in that collection assets may be employed mainly to obtain confirming evidence to the neglect of disconfirming evidence. This in turn would increase even more the bias toward remaining with early decisions.

The confirmation bias has been found in several other realistic decision contexts, including scientific inference (Mynatt, Doherty and Tweney, 1977), medical diagnosis (Kern and Doherty, 1982), and consumer information search (John, Scott and Bettman, 1986). Interestingly, a study of belief revision in auditing (Ashton and Ashton, 1987) failed to show this effect, and in fact the subjects in that study tended to be more responsive to disconfirming than to confirming evidence. The authors suggest that this may be a function of (1) the nature of the auditing task, which is often conceptualized as an evidence collection/evaluation process in which beliefs about financial statement assertions are revised on the basis of new evidence, and (2) the nature of the education and training received by auditing students. Since evidence collection and evaluation are integral parts of the situation development task in military intelligence, it may be that the nature of evidence collection and evaluation should be more heavily emphasized in intelligence training than it is currently.

There has been some debate in the literature as to whether a hypothesis-testing strategy that is based on the seeking of confirming evidence is necessarily a poor strategy (see for example, Hardin, 1980, and Tweney, Doherty and Mynatt, 1982). Klayman and Ha (1987) have recently reviewed this issue and concluded that "positive testing" (i.e., seeking confirming evidence) not only can provide useful information but in some cases is the only strategy that can reveal conclusive falsifications. The latter occurs when the hypothesis being tested is not specific enough (i.e., it

encompasses the truth), in which case it is impossible to narrow it by finding a negative instance. Thus the only possible result giving conclusive falsification is for an instance to be hypothesized as true and found to be false (i.e., positive testing). Klayman and Ha make a distinction between two meanings of the term "seeking disconfirmation," which they point out have been confused in the literature. One meaning is to examine instances that you predict will be false; the other is to examine instances you most expect to falsify rather than verify your hypothesis. Fischhoff and Beyth-Marom (1983) have also pointed out that the term "confirmation bias" has been used to describe biases in both information search and interpretation.

Unfortunately these discussions are only marginally related to the behavior found in the work by Tolcott et al. (1987) and Serfaty, et al. (1988). Klayman and Ha (1987) deal with cases in which the truth is known, and the hypothesis tester is given accurate feedback after each test; the important issue for them is which type of test is more effective. The confirmation bias found in our work on situation development by intelligence analysts was, rather, the tendency to give undue weight to positive or negative information, where the information was not actively obtained but passively received (this is not the case in collection management, of course). Furthermore, in intelligence analysis, the information received is only probabilistically related to the hypotheses and no immediate feedback is provided. As pointed out by Klayman and Ha, however, the consequences of using a positive testing strategy will depend on the nature of the task, and can lead to inefficiency or inaccuracy by overweighting some data and underweighting others, which is essentially what we found in Phase 1.

2.0 THE PHASE 2 EXPERIMENT

2.1 General Description

The Phase 1 results suggested that the types of confirmation bias found (high confidence in early judgment, overweighting of positive evidence, reluctance to change hypotheses) might be overcome through more effective use of aids (primarily graphic) and training innovations that would highlight uncertainties and encourage the active consideration of alternative hypotheses. With this in mind, a Phase 2 experiment was planned and conducted at Fort Huachuca in April 1988. The remainder of this report deals with the Phase 2 experiment.

The Phase 2 experiment employed essentially the same scenario as Phase 1, with slight modifications as described below. (The scenario and experimental materials are described in detail in Tolcott, Marvin and Lehner, 1987.) The significant difference was the introduction of several techniques aimed at encouraging the participants to assess new information in the light of several possible enemy avenues of approach, rather than concentrating their attention on confirming their initial hypothesis. Ideally it would have been preferable to test each technique separately, but this would have required a prohibitively large number of participants. For the same reason, no separate control group (a group with no interventions) was used; rather, the Phase 1 condition, which did not include the interventions, was taken as the control. As in Phase 1, the participants were mostly officers (Captains) with military intelligence specialties, taking the advanced course at the U.S. Army Intelligence Center and School (USAICS) at Fort Huachuca, Arizona. Since only 16 students were taking the course, and we had asked for 20 participants, several instructors and staff personnel were made available. As in Phase 1, the participants worked in pairs, to promote conversation between them that would reveal the rationale for their judgments. Unlike Phase 1, their conversations were not recorded.

2.2 The Interventions

The following interventions were introduced to determine if the confirmation bias could be overcome: instruction on decision biases, an enemy order of battle (EOB) graphic chart, a workbook reorganized by equipment items, and event templating instructions and graphics. These are described below.

2.2.1 Instruction on decision biases. During the introduction to the session, about three minutes was devoted to a description of several judgmental tendencies commonly found in decision research, specifically the tendency toward overconfidence in early decisions and reluctance to change, overweighting of confirming evidence and underweighting of disconfirming evidence, seeking confirming evidence, and ignoring uncertainty and base rate information. Participants were told specifically that previous work at Fort Huachuca had produced similar findings. This material was intended to represent in very brief form the type of information that might be included in a 1-2 hour instructional module in the training course.

2.2.2 Enemy Order of Battle graphic chart. One of the standard items used in situation development (and provided to the participants in both phases of this research) is an OB workbook, which contains information about enemy unit composition (subunits and equipment), disposition (including a question mark when location is unknown), strength, tactics and miscellaneous. Recall that the Phase 1 participants tended to work primarily at the maps and overlays, to ignore the uncertainties and base rates that could have been determined (or computed) from the OB workbook, and to associate units and equipments more or less intuitively from the graphic displays. In an effort to make some of the OB workbook information available to them in graphic form in Phase 2, the OB information on unit composition was displayed as coded markers adjacent to the map/overlays, in such a way that when a unit was located, the marker could be removed from the chart and placed on the map overlay in the correct position. Thus, enemy units that had not yet been located were represented by markers still on the chart, making the degree of uncertainty in enemy unit disposition very obvious. Also, unit strengths were represented graphically by degree of shading within a marker, and organization was shown with color codes. When a

marker was moved to the map overlay, its outline remained on the chart, so that the complete OB was always visible. Figure 1 shows a portion of the graphic OB chart.

2.2.3 OB-by-Equipment workbook. The OB workbook is traditionally organized by unit. Thus, if an item of equipment is detected on the battlefield, the analyst must refer in the workbook to a different page for each unit that might own the equipment to determine how many such items it owns (a base rate that would help determine ownership). This is an onerous procedure, and was rarely followed in Phase 1. To help the analyst determine ownership, an OB workbook organized by equipment rather than unit was made available. Figure 2 shows a sample page. Thus, when an unassociated item of equipment was discovered, the analyst could refer in the workbook to one page dealing with that equipment, and find a list of all units that own such an item and how many they own, and thus determine the odds that the equipment may belong to any particular unit.

2.2.4 Event template. Event templating is the procedure of plotting on an overlay the location of named areas of interest (NAIs) (e.g., road junctions, river crossings) and critical events that would be expected to occur if the enemy chose an avenue of approach (AOA). Analysts are taught, as part of the Intelligence Preparation of the Battlefield (IPB) to develop event templates for each possible AOA, to help them match intelligence reports with expected events in assessing enemy activity. In Phase 1, event templates were rarely generated; occasionally a team would identify events that would probably occur if their hypothesized AOA were correct, but they failed to do this for alternative AOAs. Event templating is a technique with procedural and graphic components that should encourage analysts to keep alternative hypotheses under active consideration. Therefore, in Phase 2 participants were given a brief explanation of event templating, immediately after the initial judgment had been made, and were offered matrices and pre-plotted event templates for their use if desired. Figure 3 illustrates the event matrix and Figure 4 the event template.

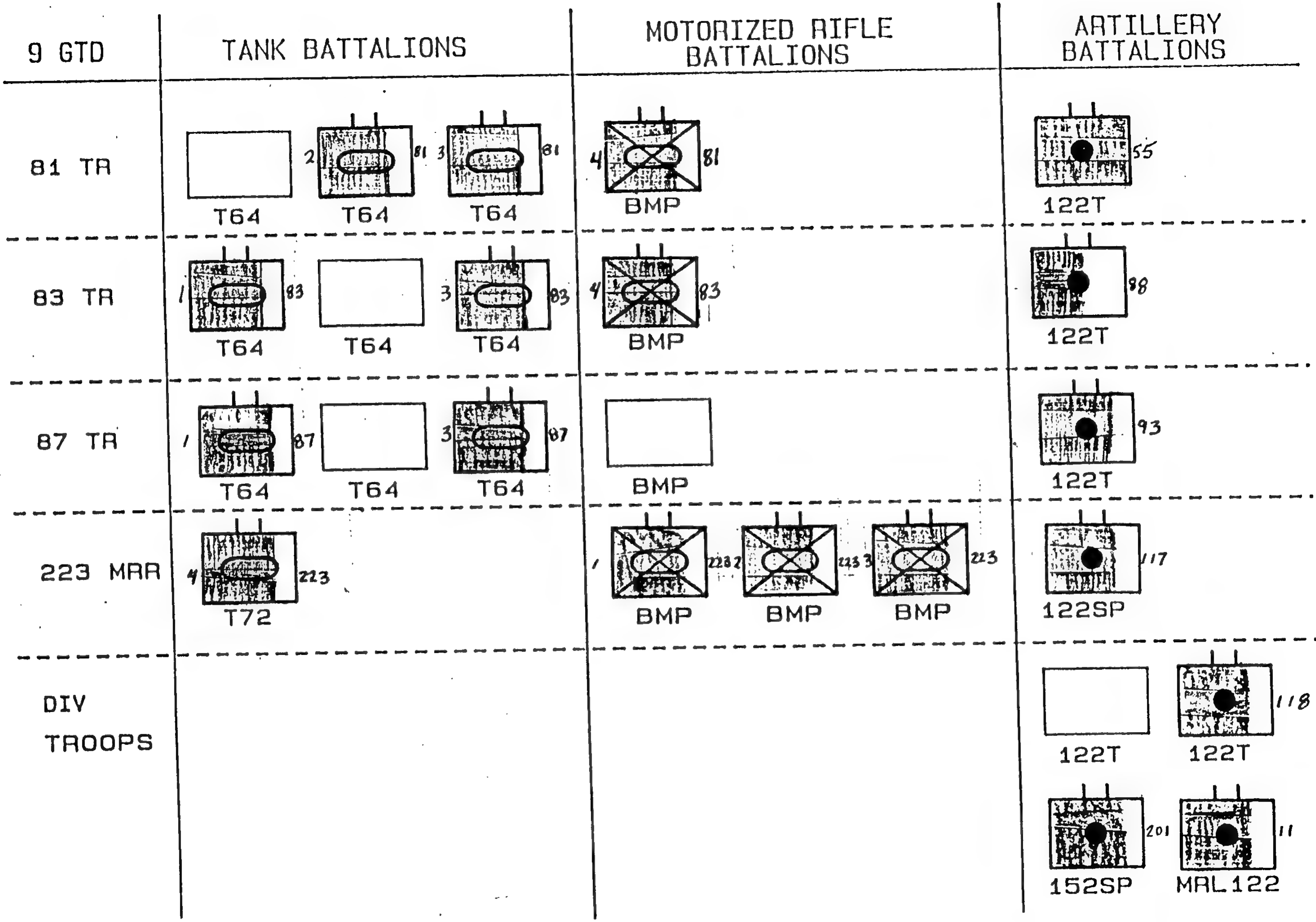

Figure 1. Portion of Graphic OB Display

EQUIPMENT CODE 59 - TANK T64

Unit Name	Division Basic Load	Division Current Estimate	Regiment Basic Load	Regiment Current Estimate	Battalion Basic Load	Battalion Current Estimate
9 GTD	282	235				
81 TR			94	80		
1-81 TBN					31	26
2-81 TBN					31	26
3-81 TBN					31	26
83 TR			94	75		
1-83 TBN					31	25
2-83 TBN					31	23
3-83 TBN					31	25
87 TR			94	80		
1-87 TBN					31	26
2-87 TBN					31	26
3-87 TBN					31	26
71 GMRD	214	135				
62 MRR			40	22		
4-62 TBN					40	22
65 MRR			40	24		
4-65 TBN					40	24
76 MRR			40	24		
4-76 TBN					40	24
11 TR			94	65		
1-11 TBN					31	20
2-11 TBN					31	22
3-11 TBN					31	22
128 MRD	40	32				
51 MRR			40	32		
4-51 TBN					40	32

Figure 2. Sample Page from OB-by-Equipment Workbook

EVENTS ANALYSIS MATRIX

Northern Avenue of Approach	Coordinates FM: NB 5050 - NB 5545 TO: NB 1855 - NB 1848
Mobility Corridor A	FM: NB 4850 TO: NB 2155

Named Area of Interest	Distance	Time	Event/Activity	Observed Time
NAI #1 NB 4753 Road junction	6 KM	20 MIN	A. RECON B. ADV GUARD	
NAI #2 NB 4655 Road junction	3 KM	10 MIN	A. RECON B. ADV GUARD	
NAI #3 NB 3859 Bridge	10 KM	30 MIN	A. RECON B. ADV GUARD C. BRIDGING ASSETS	
NAI #4 NB 2954 Road junction, city boundary	12 KM	40 MIN	A. RECON B. ADV GUARD	
NAI #5 NB 2655 Bridge	4 KM	15 MIN	A. RECON B. ADV GUARD C. RIVER CROSSING	
NAI #10 NB 2453 Road junction	4 KM	15 MIN	A. RECON B. ADV GUARD C. DEPLOY MAIN BODY	

Figure 3. Portion of Events Analysis Matrix

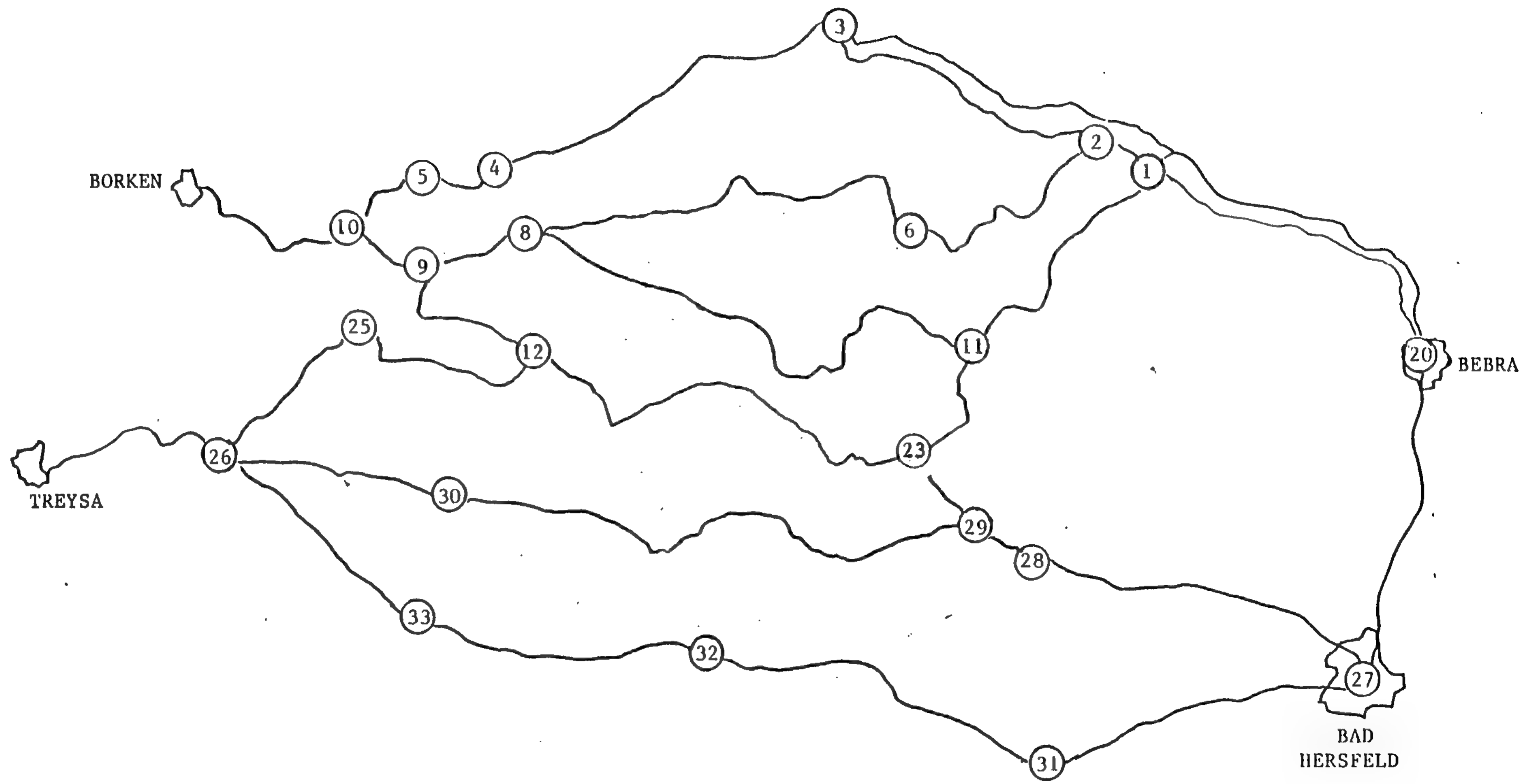

Figure 4. Event Template

2.3 Scenario Modifications

In order to correct several minor problems and inconsistencies, the problem scenario was slightly modified for the Phase 2 experiment. First, since only two out of eleven teams in Phase 1 had chosen a northern AOA initially, a northern main attack was made more likely by weakening the friendly forces on the north flank (the German corps sector) and the enemy forces in the southern sector. Second, the introduction to the scenario was expanded to include a brief discussion of the "big picture," the current situation across the entire theater front. Third, the locations of several events in the updating reports were shifted slightly to correspond more closely with the probable NAIs on the event template. It was felt that none of these modifications would distort the experimental conditions, but would simply help balance the north-south uncertainty, provide clarifying information, and be consistent with the graphic aids. Finally, only two rather than three versions of the initial scenario were used, the one favoring north and the one favoring south. Condition C of Phase 1, the balanced scenario, was omitted on the assumption that by so doing the number of cases falling into the north and south response cells would be increased.

2.4 Procedure

As stated above, the procedure was essentially the same as in Phase 1 (described in detail in Tolcott, Marvin and Lehner, 1987), with the exception of the interventions described above. The scenario was a modified version of a Central European (Fulda Gap) scenario used during training at the Command and General Staff College at Fort Leavenworth, Kansas. Combat had begun, an enemy attack was expected against the U.S. 52nd Division, and the task was to determine the direction of the enemy's main attack, initially and again after each of three update reports, and to give confidence level (on a scale from 0 to 100%), and reasons for the judgment. Ten teams of two participants each were allowed up to four hours for the task. Five teams were given an initial scenario favoring a northern attack and five a southern. The update reports were identical for all teams; each consisted of 15 items, 3 suggesting north, 3 south, and 9 neutral. After the third update, participants were asked to rate all 45 items on a five-point scale: +2 for strongly confirming their initial decision, +1 for weakly confirming it, 0 for neutral, -1 for weakly

disconfirming, and -2 for strongly disconfirming it. (This procedure had to be modified slightly for several teams who changed their decisions during the exercise, as will be described in the Results section.)

Thus, the sequence of events was as follows:

1. Introduce the session, including assurance of anonymity and information on general tendencies in decision making and results of Phase 1.
2. Read general instructions, setting the context and task.
3. Read G-3 briefing at the large-area wall map.
4. Read Division Commander's guidance at the small-area map.
5. Repeat assigned task.
6. Present and describe Intelligence Workbook and OB Workbook.
7. Present and describe OB-by-equipment Workbook.
8. Present and describe OB graphic chart.
9. Present other materials (overlays, marker pens).
10. Team works problem and gives initial estimate and confidence.
11. Review event templating procedure and offer events matrices and pre-plotted event templates.
12. Present first update report.
13. Team works problem and gives revised estimate and confidence.
14. Repeat steps 12 and 13 with second and third update reports.
15. Team rates items in the update reports.
16. Discussion of the exercise.

2.5 Subjects

Subjects consisted of 20 officers (3 1st Lieutenants, 15 Captains, 1 Major and 1 Chief Warrant Officer), all at the U.S. Army Intelligence Center and School (USAICS), Fort Huachuca, Arizona. Fourteen of the officers were about 8 weeks into a 10-week Officer's Advanced Course (OAC) qualifying them as 35D's

(tactical intelligence); 12 of these had primary speciality designations of 35E (formerly 36A, counter intelligence), 35C (image interpretation), or 35G (signal intelligence), and two were air defense officers with little or no prior intelligence experience. Four of the officers were instructors or former instructors in OAC who already were designated 35D's. One of the officers had prior enlisted experience in an intelligence speciality, 98G (voice intercept), and was currently TOC support platoon leader in a CEWI Battalion. The warrant officer had a designation of 964A (prior enlisted MOS 96B, intelligence analyst).

Most (14) of the subjects had 5-9 years of active-duty military experience, while six had from 11 to 16 years. Their experience in the intelligence field ranged from a low of 8 weeks (in the OAC course) to a high of 16 years, with the others ranging from 1-1/2 to 12 years. Fourteen of the subjects had had at least some overseas experience, but little if any operational experience doing tactical intelligence analysis.

3.0 RESULTS

In this section, the results of the Phase 2 experiment will be presented and compared with those found in Phase 1, since Phase 1 is considered the control (i.e., non-intervention) condition.

3.1 Initial Decisions and Confidence Levels

The initial information given to the participants was essentially the same, except that for five of the teams (Group N) the scenario slightly favored an enemy main attack to the North, and for the other five (Group S) it slightly favored an attack to the South. The distribution of initial estimates by group is shown in Table 3-1, which shows also the responses obtained in the Phase 1 study.

Table 3-1: Distribution of Initial Estimates, by Group

Initial Estimate	Group N	Group S	Total	Phase 1 Responses
North	4	2	6	2
Center	0	0	0	1
South	1	2	3	7
Far South	0	1	1	1

The responses obtained included north, south, and far south (an approach to the south of the divisional area). Comparison with the Phase 1 responses suggests that our attempt in Phase 2 to avoid the unbalanced distribution in favor of South (found in Phase 1) by modifying the scenario slightly, in fact tilted the balance somewhat toward North. However, a more balanced distribution was obtained. In Phase 2 there was no "Center" group as there was in Phase 1, and no Center responses were obtained. However, as in Phase 1, one team estimated that the main enemy attack would be to the south of the divisional boundary, on the grounds that the primary enemy objective was Frankfurt and that the best approach was the road network to the south of the 52nd Division.

The initial confidence level for each team is shown in Table 3-2. The first point to note is that the average confidence level in Phase 2 was 67.0, as compared with 77.3 in Phase 1. This difference is not statistically significant ($T = 1.98$, $p = .063$), but approaches significance. Secondly the average confidence of the North group was significantly higher than that of the South group (77.0 to 57.0; $T=3.36$, $p = .01$) but the confidence seems unrelated to the consistency between the scenario and the response. Thus, in the North group (for whom the scenario favored North) one of the highest confidence levels was expressed by Team 5N, who responded South. Similarly, among the South teams, the confidence levels were approximately the same regardless of whether the response was North or South. It should be noted that Team 4S, which responded Far South for the main enemy attack, gave a 40% confidence level for this response, and stated that the remaining confidence (60%) should be about equally split (30-30) between North and South for the possibility of a diversionary attack.

Table 3-2: Initial Confidence Levels

Team	Initial Estimate	Confidence (%)
1N	North	65
2N	North	85
3N	North	80
4N	North	70
5N	South	85
1S	South	65
2S	North	60
3S	South	60
4S	Far South	40
5S	North	60
	Average:	67.0
	(Phase 1 Average):	77.3

One possible reason for the difference in confidence level between the N and S teams might have been an inadvertent bias in assigning participants with more experience to one group that the other (participants were randomly assigned in the sequence in which they arrived in the exercise room). It might be hypothesized, for example, that a more experienced team would be less confident in an assessment because they were sensitive to a greater variety of factors than a less experienced team. However, an examination of the years in

service by the participants showed an average of 9.4 years for the N group and 8.2 years for the S group, hardly a significant difference.

3.2 Effects of Subsequent Information

As described earlier, all teams were given the same updating information, in three sets of 15 items each. After each set of updates, they were asked to review the situation, give an updated estimate of most likely enemy avenue of approach, their reasons, and their confidence level.

A major finding was that, unlike Phase 1 where only one team out of 11 changed their initial decision (at the last update), in Phase 2 five teams switched (one of them twice), and only five teams remained with their initial decision throughout. For purposes of data analysis, the switchers and non-switchers will be treated separately.

Table 3-3 shows the trends in confidence level for the non-switchers, both individually and averaged, and presents the overall average in Phase 1, for comparison. Team 4S maintained a confidence level of 40% that the main attack would be to the Far South, south of the division's boundary, and that there was an even chance that a diversionary attack would come in the north or south division sectors. The average for the non-switchers is given with Team 4S both included and excluded.

Table 3-3: Trends in Confidence Level (Non-Switchers)

Team	Initial Decision	Initial Confidence	1st Update	2nd Update	3rd Update
1N	North	65	70	72	74
2N	North	85	85	70	85
2S	North	60	60	50	80
1S	South	65	65	65	75
4S	Far South	40	40	40	40
	Average:	63.0	64.0	59.4	70.8
Avg. (excluding 4S):		68.8	70.0	64.2	78.5
Avg. (Phase 1):		77.3	79.6	82.2	80.0

It is clear from Table 3-3 that the confidence levels of the non-switchers in Phase 2 are generally lower than was found in Phase 1, but that the trend is for confidence to increase in response to the same new information, regardless of the initial decision, as in Phase 1. In fact, the amount of increase between initial and final confidence is even greater for this group, on average, than was found in Phase 1, whether or not Team 4S is included. Thus we may conclude that at least some of the participants exhibited the same trend of increasing confidence in an early decision, although at a generally lower confidence level, despite the interventions that were introduced in this phase.

Of perhaps greater significance, however, is that the interventions apparently had a substantial effect on half the teams, as evidenced by the fact that they changed their initial assessments during the exercise in response to the updating reports; one team in fact shifted twice. Table 3-4 presents the decisions and confidence levels for these five teams. The confidence levels in this table are always the confidence in the team's then-current decision (the data will be treated differently later).

Table 3-4: Decisions (D) and Confidence Levels (C) (Switchers)

	Initial		1st Update		2nd Update		3rd Update	
Team	D	C	D	C	D	C	D	C
5N	S	85	N	72.5	N	85	N	85
5S	N	60	N/S	50/50	S	70	S	85
3N	N	80	N	70	S	80	S	70
3S	S	60	S	60	S	60	N	65
4N	N	70	N	50	S	55	N	60
Average		71.0		60.5		70.0		73.0

The first thing to notice is that the teams that switched represent those 1) for whom the initial scenario favored both north (5N, 3N, 4N) and south (5S, 3S), and 2) who gave initial responses that were both consistent (3N, 3S, 4N) and inconsistent (5N, 5S) with their initial scenario. Secondly, it should be noted that the average initial confidence level of this group was 71.0, somewhat higher than the average of the non-switchers when Team 4S is included (63.0), and about the same as the average excluding Team 4S (68.8). Thus, we may conclude that the subsequent shifts in assessment by this group were due not to the initial scenario conditions, or to the consistency of their initial

assessment with their scenario, or with their initial confidence level. We conclude, therefore, that their shifts in assessment were due rather to how they responded to the evolving situation, or more precisely, how they responded to the items in the update reports. We might characterize them as data-driven rather than hypothesis-driven. Furthermore, since half the teams in the Phase 2 exercise were data-driven, as compared to only one out of 11 teams in Phase 1, we may infer that the interventions in Phase 2 were responsible for producing this effect.

Secondly, this group, like the non-switchers, showed an average confidence level lower than that found in Phase 1 (refer to Table 3-3); however, they did not increase their confidence as much as the non-switchers did between the first and the last assessment. The number of cases is too small for a statistical test, but Figure 5 shows the general trends.

In order to make the results more comparable to those found in Phase 1, the data should be examined another way. In Phase 1, the confidence level was always expressed in terms of confidence in the initial decision. For the one team that changed their decision from South to North after the third update, and gave a confidence level of 50-55% in their new estimate (North), the response was transformed to 45-50% - or 47.5% - confidence in the earlier estimate (South) for purposes of data analysis. This treatment assumes that the choice was binary, and that if confidence in a South estimate is C, the confidence in a North estimate would be 100-C. If the same procedure is applied to the responses of the Phase 2 teams that switched, namely, applying the formula 100-C to convert confidence level from the new to the old estimate, the data would be as presented in Table 3-5. The table indicates more clearly the trend for the switchers to show a general reduction over time in confidence in their initial estimate.

If we now consider all ten teams, and simply count the number whose confidence in their initial decision increased from the beginning to the end of the exercise, we find that number to be 3, or 30%, as compared with 7 out of 11, or 63.6%, of the teams in Phase 1. Thus, in general we may conclude that the interventions lowered the confidence level and reduced the tendency for confidence in an initial decision to rise.

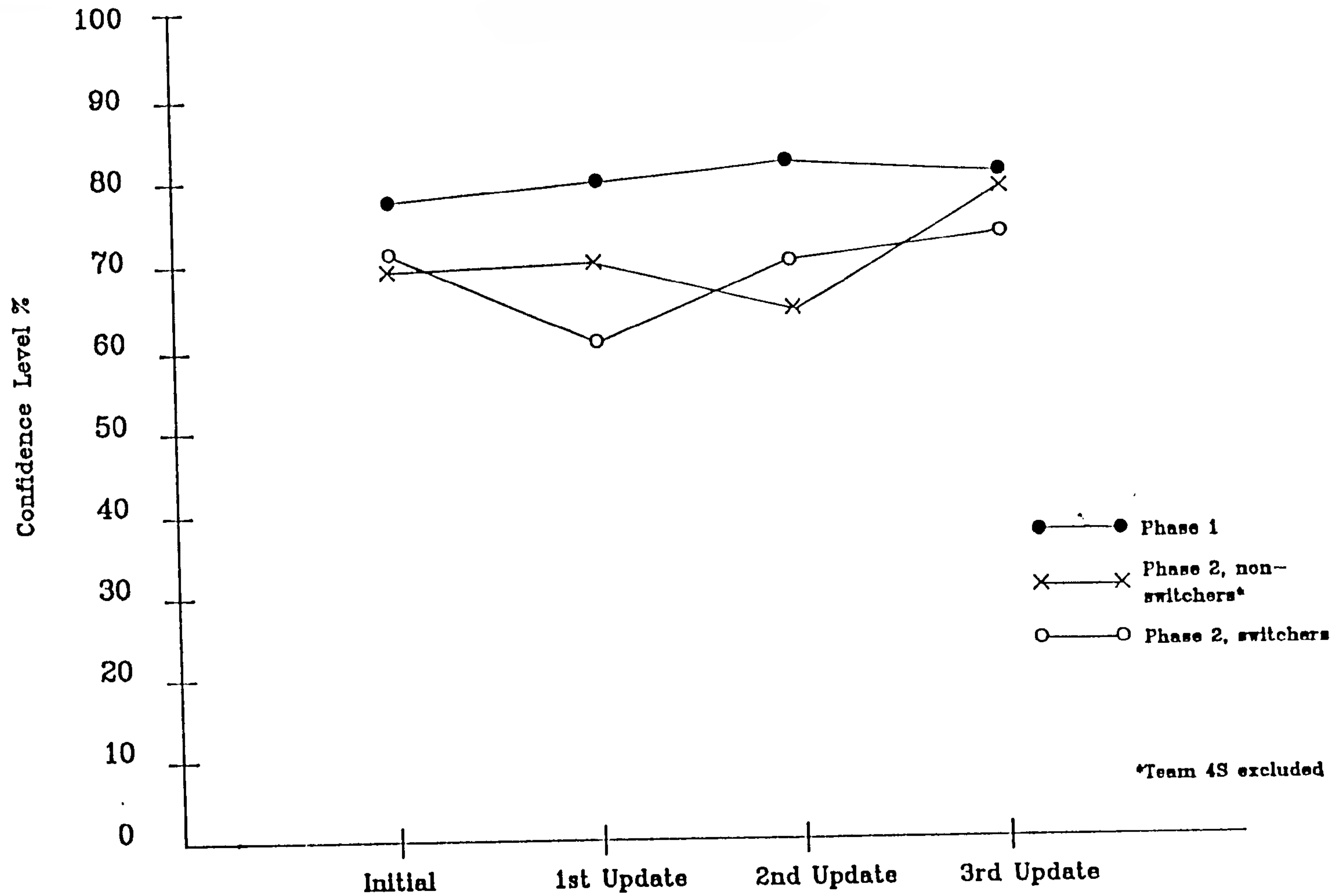

Figure 5: Trends in Average Confidence Level.

Table 3-5: Switcher's Confidence Levels Converted to Confidence in Initial Estimates

Team	Initial C	1st Update C	2nd Update C	3rd Update C
5N	85	27.5	15	15
5S	60	50	30	15
3N	80	70	20	30
3S	60	60	60	35
4N	70	50	45	60
Average:	71.0	51.5	34.0	31.0

3.3 Ratings of Information Items

After the third (and final) update, the participants were asked to review all the information items in the three update reports, and rate each item with regard to the extent to which it supported or contradicted their estimate. A rating of +2 indicated strong support; +1 weak support; 0 neutral; -1 weak contradiction; and -2 strong contradiction. In the Phase 1 exercise, these ratings were all given with respect to the initial estimate. Since only one team in Phase 1 switched estimates, and the switch occurred after the last update, it was natural for the teams to rate the items with respect to their initial estimates.

Since half the teams in Phase 2 changed their initial estimate, a different procedure was used. One of the switching teams changed after the final update, and they were asked to rate the items with respect to their initial estimate, as was done by the comparable team in Phase 1. The other switchers, who changed their estimates earlier in the exercise, after the first or second update, were asked to rate the items with respect to their newer estimate. An exception was made in the case of the double switcher - the team that gave an initial estimate of North, switched to South at the 2nd update, and switched back to North at the 3rd update; this team gave ratings with respect to their initial decision, North.

It should be added that an intervention seriously considered was to have the participants rate each item with respect to both a North and a South estimate at the time they received the reports. This procedure might have sensitized them to the uncertainties and encouraged them to keep both hypotheses in mind.

Indeed, Einhorn and Hogarth (1987) suggest that inertia is greater when the updated belief is called for after all the evidence has been presented, than if new responses are requested with each piece of new evidence. We did not adopt this procedure because we wanted to keep the responses (dependent variables) the same as in Phase 1, for comparison purposes.

Table 3-6 shows the average arithmetical total score given to each update report (15 items) by teams grouped according to their initial decision, for those teams whose ratings were given with respect to their initial decision. Team 4S (far South) was excluded.

Table 3-6: Average Arithmetical Total Score for Each Update Report (Ratings with reference to initial decision)

Initial Decision	Average Score		
	1st Update	2nd Update	3rd Update
N (4 cases)	+3	-1	+7
S (2 cases)	+3	+2.5	-4

In Phase 1, there was a clear tendency to regard the update reports as supportive of the initial decision, whatever that decision was; in fact, there were no negative average totals. The Phase 2 results, on the other hand, suggest a much closer attention being paid to the diagnosticity of each item, resulting in two negative average totals. It should be noted that a negative total score for a report was not necessarily associated with a change in estimate, since the estimates were based not only on the update reports but also on certain constant factors (such as initial scenario, intelligence and enemy OB, terrain, etc.). But these results show that the interventions did reduce the tendency to regard new information as supportive of an early decision.

Table 3-7 gives the individual and average data for the three teams who changed their estimate early and rated the items with respect to the new estimate.

Table 3-7: Ratings (R) of Early Switchers, with respect to new Decision (D)

Team	Initial Decision	1st Update R	1st Update D	2nd Update R	2nd Update D	3rd Update R	3rd Update D
5S	N	+4	50/50	+6	S	+2	S
3N	N	0	N	+5	S	+1	S
5N	S	+3	N	-3	N	+4	N
	Average:	+2.3		+2.7		+2.3	

These results are difficult to interpret. The tendency of this group is to find the update reports supportive, especially for the time at which a change of estimate was made. On the average, the total scores are relatively low, although positive. These results suggest that the decisions of this group were more heavily influenced by the update reports than were the relatively more stable teams represented in Table 3-6.

Chi-square analyses were conducted to compare the obtained with the expected distributions in each phase, as well as between the distributions obtained in Phases 1 and 2. For this purpose the negative ratings (-2 and -1) were combined, as were the positive ratings (+1 and +2), and the neutral (0) ratings discarded. Table 3-8 shows the distributions. For the intra-phase analyses the expected distribution for the null hypothesis (no bias) is an equal number of negative and positive ratings. The results show that in Phase 1 the distribution is significantly different from the expected ($df = 1$, $p < .005$), indicating a definite bias. In Phase 2 the obtained distribution is also significantly different from the expected ($df = 1$, $.01 < p < .025$), indicating that a tendency to regard evidence as supportive still exists.

For the inter-phase comparison, interest centers on whether the interventions had an effect, therefore the Phase 1 (control group) distribution is regarded as the expected. The results show that the Phase 2 distribution is significantly different from that of Phase 1 ($df = 1$, $p < .005$), in the direction of a more equal distribution of positive and negative ratings. Therefore we may conclude that the interventions did indeed have the effect of reducing, but not eliminating, the confirmation bias.

Table 3-8: Distribution of Positive and Negative Ratings and Chi-Square Results

		Direction of Rating -	+	
Phase 1	Expected	103	103	df = 1, $p < .005$
	Observed	50	156	
Phase 2	Expected	84	84	df = 1, $.01 < p < .025$
	Observed	69	99	
	Phase 1	50	156	df = 1, $p < .005$
	Phase 2	69	94	

Another possible reason for the differences between Phase 1 and Phase 2 is that the Phase 2 participants might have been a more experienced group (it will be recalled that the Phase 2 participants included several instructors). To test this possibility, the mean years of Army experience of the two groups was compared. The mean of the Phase 1 group was 6.9, and for the Phase 2 group, 8.8, showing slightly more experience in Phase 2. However, this difference was not statistically significant ($P = .11$). A similar comparison was made between the Phase 2 switchers ($M = 7.0$) and non-switchers ($M = 9.2$). This difference is in the unexpected direction, since one would expect the switchers to be more experienced, but again the difference is not significant ($P = .67$).

One final possibility is that the difference was due to the fact that the Phase 2 student participants were further into the school course (8-9 weeks) than were the Phase 1 participants (3-4 weeks). Although this possibility cannot be completely ruled out, our information about the course, from both the school curriculum and discussions with the participants, indicates that the instruction emphasizes procedural rather than inferential processes, and does not include exercises that illustrate an evolving situation. A supplementary text, dealing with decision making and describing typical biases found in the research literature, is available to the students, but there was no evidence that any of the students had referred to it. In any event, this

text is more theoretical than practical, since it does not relate the biases to examples taken from intelligence analysis.

It seems safe therefore to conclude that, despite the lack of precise control over all possible relevant variables, the differences found between Phases 1 and 2 were due to the interventions introduced in Phase 2.

3.4 Discussion of Information Used

As in Phase 1, the participants worked primarily at the map displays, after reviewing the guidance and the materials in the Intelligence and OB Workbooks. Although several types of new material (OB-by-equipment workbook, OB graphic chart, event matrix and event template) were made available as experimental interventions (i.e., they were not available to the Phase 1 teams), the materials were not used to an equal extent.

By far the most frequently used aid was the graphic display of enemy OB, and the participants all commented positively on its usefulness. The benefits of this aid were that (1) it showed at a glance both the amount of uncertainty (number of unlocated enemy units) and the base rates (number of units belonging to parent units), (2) it showed additional information in coded form (unit strengths and organization), and (3) it enabled the participants to remove counters of located units and place them on the map, thereby integrating the counters into the display that they were working at. Several teams suggested that the unit counters should also show major items of equipment owned by the units. The OB-by-equipment workbook was used occasionally by some of the participants to identify units that owned an item of equipment and determine the number of items owned by each unit. However, only two teams thought this workbook was useful, and it was not used as frequently as had been anticipated. One team remarked that they were more comfortable using materials with which they were familiar. In retrospect, it appears that this book should have been tabbed for easier access, and that more time should have been spent explaining its use. The concept of entering a data base by equipment item to determine possible owning units still merits consideration; with a computerized OB data base this concept would be easy to implement.

The event matrix and template were largely ignored; only two teams referred to it, and only one thought it was helpful. Comments suggested that they had not received sufficient training in event templating procedures, that it took too long to prepare it (that is, to anticipate and plot enemy activity at various future times under several possible enemy courses of action), and that the terrain analysis provided the needed information. Even participants who were instructors (and reasonably familiar with event templating procedures), commented that they were not satisfactory. Event templating is one of the few, if not the only, standard procedures that explicitly forces the situation developer to consider alternative enemy actions. These findings suggest that if event templating is regarded as important, the procedures should be simplified or more training time should be devoted to them, or both.

4.0 CONCLUSIONS

4.1 Discussion of Findings

The major finding was that the interventions in Phase 2 caused the participants to become more sensitive to the implications of the data they were receiving and less influenced by their initial estimates of the situation. Their confidence in their initial judgments tended to be lower than was found in Phase 1, and five of the teams (50%) changed their initial estimates at least once during the course of the exercise, as compared with one team in Phase 1. The confidence of the teams that switched remained lower than the teams in Phase 1, but the confidence of the teams that maintained their initial estimate rose to about the same level as the Phase 1 teams by the end of the exercise.

With regard to weighting the evidence as confirming or disconfirming, the interventions significantly reduced the tendency to overweight confirming evidence, although the tendency was not eliminated.

As in Phase 1, little if any attention was paid to negative indicators (i.e., predicted events that did not occur). Nor did the participants bother to predict events that would occur with alternative enemy actions, despite the suggestion that available event templates might be useful in this regard. Rather, the lowered confidence of the switchers and their willingness to change their estimates appeared to result from their better awareness of the ambiguity inherent in positive indicators. As an example of this, one of the update items reported the capture of an enemy regimental commander while performing forward reconnaissance in the southern sector. In Phase 1, most of the teams regarded this a strong support for an enemy attack in the south, because of the presence of a high-level enemy officer so far forward. Many of the Phase 2 teams took a more skeptical view of this piece of evidence, and one team remarked that it was just as likely that a high-level enemy officer was doing the same thing in the northern sector, except that he had not been captured. This type of thinking was more common in Phase 2 than Phase 1, and was reflected in the significantly fewer strong positive ratings of the information items.

As in Phase 1, graphic information was preferred to tabular; participants plotted update information on the map overlays and drew conclusions from these graphic displays. The ease with which the OB chart's unit counters could be placed on the overlays was a significant factor in their acceptance and use by all the teams. This display, and to a smaller degree the OB-by-equipment workbook, seemed to reduce the tendency to ignore base rates which was prevalent in Phase 1. There is no question that the graphic presentation of OB data increased the participants' awareness of the uncertainty due to the large number of enemy units that had not yet been located, and therefore may have contributed to the lowered confidence and reduced confirmation bias. The OB-by-equipment workbook might have played a more significant role if its design had been improved by index tabs to facilitate access.

It is likely that the confirmation bias would have been reduced even more by requiring the participants to assess the diagnostic strength of each update item as it was received, rather than at the end of the exercise. This procedure would perhaps make more obvious the extent to which certain evidence might support more than one hypothesis, and make the participants more aware of any tendency to overweight supporting evidence. Although such a procedure would probably be too time-consuming to be followed during actual combat, its practice during training might significantly improve the inference process during operations.

Finally, the findings have possible implications for another aspect of intelligence analysis, specifically, the collection management process. This is the process of identifying additional intelligence information needed to test hypotheses about enemy intent (Priority Information Requirements, or PIR), and allocating resources to obtain it. A confirmation bias in this process might lead to expenditure of a disproportionate amount of resources to seek information that would confirm a strongly held hypothesis, while neglecting to seek information that might disconfirm it. Investigation of hypothesis testing behavior in the context of collection management appears to be a fruitful next step in this research.

4.2 Recommendations

The results of this research point clearly to the importance of 1) graphic aids to help the inference process that is the heart of tactical intelligence analysis on the battlefield, and 2) indoctrination during training to familiarize students with the common tendencies towards overconfidence and the overweighting of confirming evidence. Practice in situation development during evolving scenarios, and in judging the diagnosticity of incoming intelligence reports, might also contribute to improved analytical performance.

The graphic aids need not be implemented in computerized form; participants in this research were generally skeptical of the value of computer aids on the battlefield at levels of division and below. However, the aids recommended herein could easily be incorporated into computerized systems if policy decisions are made to develop such systems for operational use. Furthermore, computerized versions might be useful during training to illustrate the concepts involved.

4.2.1 Graphic aids. The concept of a graphic OB display should be improved and implemented. In its manual form, the major piece of equipment in each unit should be annotated on the unit counter, and the unit designation should be shown on the silhouette that remains after the counter is removed. This aid could easily be incorporated into a computerized system, as a supplement to a computerized map of the battlefield. In such a form, it would be possible to provide some indication of the temporal aspects of the evolving situation (for example, showing on each unit the time (date-time group) at which it was located, and showing predicted position of enemy units at various future times).

This concept could be expanded to become a computerized event template, by showing predicted events for various alternative enemy courses of action, a useful way to encourage constant awareness of alternatives but one that is apparently difficult to implement manually. However, more experimentation is needed to determine exactly what features should be included in such an aid.

The OB-by-equipment workbook appears to be a potentially useful way to provide base rate information, but it was not used frequently enough during the exercise to justify its development. In its manual form, it should be improved by the addition of index tabs and evaluated by instructors at USAICS. In a computerized enemy OB data base it would be easy to provide access to the data by equipment type, and if a computerized version is being developed, this feature should be added.

4.2.2 Training concepts. The few minutes spent informing the participants about common tendencies toward overconfidence and the confirmation bias apparently had an effect. It is recommended that at least one hour be devoted to this topic during the Officers Advanced Course at USAICS, with specific examples of how these biases might be manifested during situation development in an evolving battlefield scenario.

Furthermore, a training game/simulation should be developed specifically to illustrate an evolving scenario and to provide practice in assessing the diagnosticity of new information as it is received. This training module could be used at USAICS as well as in the field for skill maintenance.

4.2.3 Additional research. Research should be undertaken to investigate the extent to which the confirmation bias is found in hypothesis testing, in the context of developing information requirements for collection management.

REFERENCES

Ashton, A.H. and Ashton, R.H. *Sequential belief revision in auditing* (Working Paper). Durham, NC: Duke University, August 1987.

Einhorn, H.J. Learning from experience and suboptimal rules in decision making. In T.S. Wallsten (Ed.), *Cognitive processes in choice and decision behavior*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1980.

Einhorn, H.J., and Hogarth, R.M. Confidence in judgment: Persistence in the illusion of validity. *Psychological Review*, *85*, 395-416, 1978.

Einhorn, H.J., and Hogarth, R.M. *Adaptation and inertia in belief updating: The contrast-inertia model*. Chicago, IL: The University of Chicago, October 1987.

Fischhoff, B. and Beyth-Marom, R. Hypothesis evaluation from a Bayesian perspective. *Psychological Review*, *90*, 239-260, 1983.

Gettys, C.F., and Fisher, S.D. Hypothesis plausibility and hypothesis generation. *Organizational Behavior and Human Performance*, 1979, *24*, 93-110.

Gettys, C.F., Manning, C., and Casey, J.T. *An evaluation of human act generation performance* (Tech. Rep. 15-8-81). Norman, OK: University of Oklahoma, 1981.

Hardin, C.L. Rationality and disconfirmation. *Social Studies of Science*, *10*, 509-514, 1980.

John, D.R., Scott, C.A., and Bettman, J.R. Sampling data for covariation assessment: The effect of prior beliefs on search patterns. *Journal of Consumer Research*, *13*, 38-47, June 1986.

Lopes, L.L. *Averaging rules and adjustment processes: The role of averaging in inference* (T.R.). Madison, WI: Wisconsin Human Information Processing Program (WHIPP 13), 1981.

Kern, L. and Doherty, M.E. "Pseudodiagnosticity" in an idealized medical problem-solving environment. *Journal of Medical Education*, *57*, 100-104, February 1982.

Klayman, J. and Ha, Y-W. Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*, 211-228, 1987.

Mynatt, C.R., Doherty, M.E., and Tweney, R.D. Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, *29*, 85-95, 1977.

Serfaty, D., Entin, E.E., Shaw, J.J., Tenney, R.R., and Kelley, M.C. *Planning with uncertain and conflicting information: Experiment report (NPS-3)* (Draft Technical Report 392). Burlington, MA: Alphatech, Inc., July 1988.

Tolcott, M.A., Marvin, F.F., Lehner, P.E. *Effects of early decisions on later judgments in an evolving situation* (Technical Report 87-10). Reston, VA: Decision Science Consortium, Inc., July 1987.

Tweney, R.D., Doherty, M.E., and Mynatt, C.R. Rationality and disconfirmation: Further evidence. *Social Studies of Science*, *12*, 435-441, 1982.